Modelling and Prediction of Athletic Readiness based on Training Load

Shashvi Shah, Jeel Patel, Dev Dave, Riya Sudani, Yashaswi Patel

School of Engineering and Applied Sciences
Ahmedabad University
Email: {shashvi.s3, jeel.p1, dev.d3, riya.s3, yashaswi.p}@ahduni.edu.in

***Abstract*—Sports performance is affected by various factors such as training load, recovery, and external commitments. In basketball at the collegiate level, the players tend to be fatigued because of regular games, travel, and intense training, which can be detrimental to their readiness and performance. This research seeks to investigate the effect of training load on sports readiness and create a predictive model for RSImod (Readiness Score Index modified). The data consist of training metrics such as workload, sleep, and previous performance. Machine learning models are utilized to predict RSImod with an emphasis on explainable AI (xAI) methods to explain model choice. The outcomes of this study offer actionable advice for training optimization, such that athletes are at peak performance levels while avoiding risk of injury.**

***Index Terms*—Athletic readiness, sleep, recovery, RSI, feature selection, XGBoost, SHAP, RFE, machine learning, regression, data preprocessing, sports analytics.**

I. Introduction

THE Athletic readiness is important to maintain optimal performance and prevent injuries in college basketball players. The rigorous scheduling of their activities, such as daily games, travel, practice sessions, and studies, tends to cause fatigue both physically and mentally. Proper management of training load is essential since too much or too little training will adversely affect an athlete's performance and recovery. The capacity to accurately estimate and forecast levels of readiness can contribute to maximizing training approaches, mitigating injury possibilities, as well as overall athletic efficacy.

This research centers on modeling and forecasting RSImod (Readiness Score Index modified), a central readiness indicator, using multiple training and physiological measures. With data-driven machine learning methods, the research investigates the interaction between training load and RSImod, uses predictive models, and employs explainable AI (xAI) to explain outcomes. The analysis also seeks to identify patterns through clustering methods, allowing for personalized training advice. The results of this research will inform data-driven decision-making in athlete performance management, ultimately enhancing readiness and minimizing the dangers of overtraining or undertraining.

Dataset: The dataset consists of 3,111 instances gathered from 16 basketball players over a period of six months (September 6, 2021 – March 7, 2022). The dataset consists of 35 attributes of sleep, recovery, and sports performance. Key values are heart rate variability (HRV), resting heart rate (RHR), sleep quality, respiratory rate, and other recovery values. The Reactive Strength Index (RSI) is the target variable, which indicates an athlete's physical preparedness.

II. Methodology

The methodology adopted in this study follows a structured process to handle the dataset, investigate training load effects on athlete preparedness, and create a predictive model for RSI. The major steps taken are as follows:

A. Data Processing

The dataset underwent several preprocessing steps to ensure data quality and enhance model performance. Missing values were handled using an Iterative Imputer with an XGBoost Regressor to ensure accurate estimations. Non-numeric columns were removed to maintain numerical consistency and prevent potential issues during model training. To standardize numerical features and bring them to a common scale, StandardScaler was applied. Additionally, a correlation heatmap was generated to visualize relationships between features, allowing the identification and removal of redundant attributes, thereby optimizing the dataset for improved predictive accuracy.

B. Feature Selection

To optimize model efficiency and reduce overfitting risks, Recursive Feature Elimination (RFE) was applied using XGBoost to select the top 18 features. Additionally, SHAP (SHapley Additive Explanations) Analysis was conducted to determine feature importance and identify the top 12 most impactful features.

C. Model Training and Evaluation

Your request is already well-structured in paragraph form. However, here's a refined version for better readability:

Three machine learning models were trained and evaluated to predict RSI, each serving a distinct purpose. Linear Regression was used as a baseline model to capture linear relationships between features and RSI. Random Forest Regressor was implemented to account for complex, non-linear interactions in the dataset. Finally, XGBoost Regressor, a powerful gradient boosting algorithm, was employed to optimize accuracy by

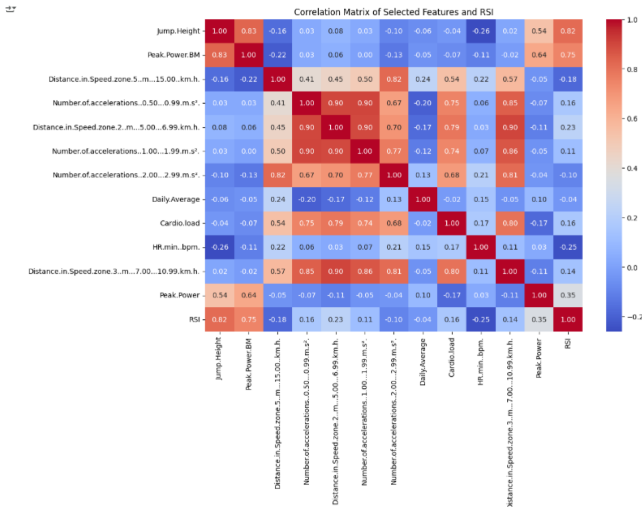

Fig. 1. Correlation Heatmap

leveraging advanced boosting techniques. The performance of these models was assessed using key evaluation metrics, ensuring a comprehensive comparison of their predictive capabilities.

D. *Feature Importance Analysis*

Feature importance was analyzed using SHAP Values, which interpret the contributions of different features to RSI, and Permutation Importance, which measures model performance variation when features were permuted.

III. Results

Table I The research compared Random Forest, XGBoost, and Linear Regression for RSI prediction, judging them based on R² Score, MAE, and MSE. XGBoost performed better with an R² score of 0.95, Random Forest with a score of 0.94, and Linear Regression scored worst at 0.80. SHAP analysis validated that incorporating only the best 12 features still had high accuracy.

A. *Key Performance Metrics*

The models were evaluated based on R², MAE, and MSE. XGBoost outperformed the other models:

TABLE I
Performance Comparison of Different Models for RSI Prediction

Model	R^2 **Score**	**MAE**	**MSE**
Linear Regression	0.80	High	High
Random Forest	0.94	Low	Low
XGBoost	0.95	Lowest	Lowest

B. *Shap Analysis*

SHAP analysis validated that using only the top 12 features maintained high accuracy. Key factors influencing RSI were Training Load (most significant feature), Sleep Duration, Recovery Index, Fatigue Level, and Heart Rate Variability (HRV).

C. *Retraining with Top 12 Features*

An optimized XGBoost model was retrained using only the top 12 features, achieving an improved R² score of 0.96. A correlation heatmap of the top 12 selected features was generated to further analyze their relationships with RSI.

IV. Discussions

The findings demonstrates that training load, recovery, and sleep quality significantly impact RSI. Higher training loads with inadequate recovery result in lower RSI values, increasing fatigue, while balanced training schedules with proper rest lead to higher RSI values, enhancing athlete readiness. XGBoost proved to be the most effective model for capturing non-linear interactions, surpassing Linear Regression and Random Forest. Additionally, SHAP analysis offered actionable insights for optimizing training regimens.

V. Conclusion

This study successfully developed a predictive model for RSI, emphasizing the impact of training load, recovery, and sleep quality on athlete readiness. The results confirmed XGBoost as the most accurate model for RSI prediction, with SHAP-based explainability validating the importance of key features. Additionally, optimized feature selection maintained high accuracy while reducing the number of predictors, enhancing model efficiency.

VI. Future Work

Integration of real-time physiological data from wearable devices can enhance dynamic assessments. Expanding the dataset to incorporate factors such as nutrition, hydration, and psychological stress could further refine predictions. Additionally, developing a real-time monitoring system would aid coaches in optimizing training regimens effectively.

By leveraging machine learning and explainable AI, this study enhances athlete performance while mitigating overtraining risks through data-driven insights.

References

[1] H. Jiang, T. Xu, Y. Liu, J. Zhou, L. Wang, and S. Yang, "Self-explaining hierarchical model for fatigue monitoring and prediction in basketball," Sci. Rep., vol. 13, no. 1, article no. 11234, 2023.

[2] J. A. Smith, P. R. Rodriguez, M. T. Johnson, and K. W. Lee, "Athletic signature: Predicting the next game lineup in collegiate basketball," J. Sports Sci. Med., vol. 21, no. 4, pp. 987-995, 2022.

[3] Y. Chen, Z. Li, T. Wang, and X. Zhao, "A hybrid approach for interpretable game performance prediction in basketball," IEEE Trans. Comput. Soc. Syst., vol. 8, no. 2, pp. 234-245, 2021.

[4] S. Senbel, S. Sharma, M. S. Raval, C. Taber, J. Nolan, N. S. Artan, et al., "Impact of sleep and training on game performance and injury in Division-I women's basketball amidst the pandemic," IEEE Access, vol. 10, pp. 15516-15527, 2022.